Parallel Sysplex Best Practices

Dan St.Cyr
Software Consultant
d.stcyr@streamfoundry.com

Share - Boston - Session 7519

Tuesday, August 3, 2010, 3:00 PM, Room 313

Dan St.Cyr

- NET/ NYNEX/Bell Atlantic/Verizon
  - 1974→1978 - Programming
  - 1978→2003 - MVS Operating System Support
    - Early customer implementation of Parallel Sysplex
    - Joint Customer Study with IBM for IMS Datasharing
    - Migrated all datacenters/systems to Parallel Sysplex
- StreamFoundry, Inc.
  - 2007→now – Consultant
    - Parallel Sysplex Health Check Reviews/updates

Abstract

A successful parallel sysplex is one that addresses the operability, availability and recoverability of a multi-z/OS environment, within available resources, while operating with “best practices” – resulting in optimum performance.

These “best practices” should periodically be reviewed to ensure the parallel sysplex is operating with minimal overhead while providing peak performance and optimal return, given the resources available.

Topics

- Coupling Facilitity - SA vs ICF implications
- Couple Datasets (CDS) - formatting/placement
- Structures – definitions/placement/features
- Signaling – setup/tuning
- System Logger – exploiting - offloading/duplexing
- Availability – COUPLExx parameters, SFM, ARM
- IBM exploiters
- RMF Reports
- Publications

Coupling Facilities

- **Hardware**
  - Single/multiple CFs (SPOF) – 3 for ultimate availability
  - **Standalone** vs **ICF**
    - Standalone or ICF w/duplexing for data sharing (failure independence)
    - ICF's fine for resource sharing (OPERLOG, LOGREC, GRS, RACF, etc)
  - Dedicated CPs (no shared CPs unless backup CF w/Dynamic CF Dispatching - **DYNDISP**)
  - Battery backup **(IBF)** or **UPS** for non-volatility – affects logger duplexing
    - With UPS, change with control code **MODE NONVOLATILE** command
  - Storage sufficient for failover
  - CF Links
    - CF to CF links for Structure Duplexing (enhanced w/CFLEVEL 16)
    - Redundant CF Channels for availability
  - D CF
- **Software (CFLEVEL of CFCC)**
  - http://www.ibm.com/systems/z/advantages/pso/cftable.html shows current levels by server model along with enhancements within each level.
  - D CF or RMF SYSRPTS(CF)

Couple Datasets

- Placement/isolation considerations:
  - Each primary/alternate pair on physically separate volumes/subsystems/controllers
  - Use high performance DASD (DFW) for large sysplex
  - Sysplex primary on different volume than CFRM primary
  - Split up the rest across minimum 2 volumes
  - Do not allow others dataset allocations on volumes – fill remaining space
  - Consider having spare CDS's available in the event of a switch
  - Reformat new Sysplex and CFRM CDS's for DR rather than dump/restore or "mirroring".

Couple Datasets (cont'd)

- Formatting:
  - Policy CDS MAXSYSTEM values should match Sysplex CDS MAXSYSTEM
  - Format alternates with ITEM's of equal value (or greater) to the corresponding primary – PSWITCH requires an alternate of >= size.
  - Define reasonable values for growth – you can increase later, but you can't easily decrease.
  - Format primary/alternate pairs using same format version
    - Reformat using new z/OS level after all systems in sysplex are at new level
  - D XCF,COUPLE,TYPE= to see currently formatted values along with peak values for MAXGROUP & MAXMEMBER

Structures

- Placement across CFs – consider size, functionality, activity
- See "Setting Up A Sysplex" Chapter 4 - *Table 7:* ***Coupling Facility Structures in z/OS for IBM Products*** - for structure type/supported functions (duplex,rebuild,alter)/documentation
- Sizing (CFSIZER) - http://www.ibm.com/systems/support/z/cfsizer/ - includes HELP
- Features to consider/review:
  - Structure full monitoring/Automatic alter (IXC585E/IXC586I):
    - **FULLTHRESHOLD(%)** – default=80 – 0 disables both FULLTHRESHOLD & ALLOWAUTOALT
    - **ALLOWAUTOALT(YES/NO)** – must be allowed by application
  - **DUPLEX(DISABLED/ALLOWED/ENABLED)** – to maintain a duplex copy of structure
    - Value (quicker failure recovery vs. rebuild) worth the overhead cost?
    - Needed due to lack of failure independence?
    - See "Setting Up A Sysplex" – Chapter 4: **An Installation Guide to Duplexing Rebuild** for requirements/implications
  - **REBUILDPERCENT**(1) in each structure for % lost connectivity initiating rebuild with SFM active
    - Rebuild happens on lost connectivity regardless of REBUILDPERCENT without SFM active
    - Does not apply to signaling structures (must be supported by application)

Signaling

- Use a combination of CF structures (bi-directional) and CTCs (uni-directional) for availability
- Initially consider:
  - Minimum 3 CF structures with CLASSLEN=956, 16316 and 62464 and GROUP(UNDESIG)
  - Avoid manually assigning XCF groups – allow MVS to choose paths
  - Supplement 1K (and maybe 16K) classes with CTC pairs – at least 4 CTC pairs from & to each system
- Define structures using minsize/size parameter to allow for SETXCF START,ALTER
- Tune using RMF XCF Activity Report

System Logger

- CF (multi-system) or DASDONLY (single-system) logstreams
- Minimally consider implementing:
  - OPERLOG
  - LOGREC
  - SMF (z/OS 1.9)
- Offloading – based on CF structure (or staging dataset) thresholds (**HIGHOFFLOAD(80)/LOWOFFLOAD(0)**) – size offload datasets via (**LS_SIZE**)
- Other IBM Logger Exploiters:
  - IMS Common Queue
  - CICS log manger
  - APPC/MVS
  - RRS (Resource Recovery Services)
- Duplexing option for 'staging' datasets for interim data recovery (rather than local buffers, eliminating a SPOF)
  - **STG_DUPLEX (YES/NO)**
  - **DUPLEXMODE (COND/UNCOND/DRXRC)** – COND - depending on CF 'volatility' and structure 'system managed duplexing'

Availability

- COUPLExx parameters – (D XCF,COUPLE)
  - INTERVAL
  - OPNOTIFY
  - CLEANUP
  - RETRY
- Sysplex Failure Management (SFM)
  (to minimize downtime - allowing SYSPLEX-management of outages)
- Automatic Restart Management (ARM)

Availability
COUPLE parameters

INTERVAL - the failure detection interval (in seconds) to determine a "status update missing" – normally based on legitimate spin loop time.

- Default computation: spinfdi = (N+1)*SpinTime + 5
  Based on EXSPATxx specifications – where:
  - N is the number of excessive spin recovery actions
    - » Default=1 through z/OS 1.10
    - » Default=3 effective with z/OS 1.11
  - +1 indicates the implicit SPIN action
  - SpinTime is the excessive spin loop timeout interval
    - » default is 10 seconds (basic)/40 seconds (LPAR)

- Resulting computations:
  Through z/OS 1.10: 25 seconds (Basic); 85 seconds (LPAR)
  As of z/OS 1.11: 45 seconds (Basic); 165 seconds (LPAR)

Recommend not coding INTERVAL - can be changed after IPL with SETXCF COUPLE,INTERVAL=

Availability
Couple parameters (cont'd)

OPNOTIFY – time value between "status update missing" and IXC402D.

IXC402D *sysname* **LAST OPERATIVE AT** *hh:mm:ss*. **REPLY DOWN AFTER SYSTEM RESET OR INTERVAL=SSSSS TO SET A REPROMPT TIME.**

- Must be >= failure detection interval
- Default is INTERVAL + 3 seconds
- SFM "ISOLATETIME" consideration
- Can be changed after IPL with SETXCF COUPLE,OPNOTIFY=

Availability
Couple parameters (cont'd)

CLEANUP – during system removal from sysplex, the max amount of time allowed for XCF group members to perform cleanup prior to being put into wait state

- Default = 15 seconds
- Can be changed after IPL with SETXCF COUPLE,CLEANUP=

RETRY – tolerance level for stopping attempted recoveries of a failed signaling path

- Default=10
- Code lower than 10 to stop path retries sooner
- Code higher than 10 to stop path retries later
- Can also be set on individual PATH statements
- Can be changed after IPL with SETXCF COUPLE,RETRY=

Availability
Couple parameters (cont'd)

New in z/OS 1.11 – XCF can adjust to an "effective" FDI – being the longer of COUPLExx-specified interval or the computed interval based on spin time.

D XCF,COUPLE

z/OS 1.10:

INTERVAL	OPNOTIFY	MAXMSG	CLEANUP	RETRY	CLASSLEN
85	88	2000	15	10	956

z/OS 1.11:

INTERVAL	OPNOTIFY	MAXMSG	CLEANUP	RETRY	CLASSLEN
165	168	2000	15	10	956

DEFAULT USER INTERVAL: 165 (or SETXCF USER INTERVAL: 100)
DERIVED SPIN INTERVAL: 165
DEFAULT USER OPNOTIFY: + 3

USERINTERVAL option in COUPLExx to force the user-specified INTERVAL value to be the effective failure detection value, even though it is smaller than the spin failure detection interval.

Availability
Sysplex Failure Management

Recommendation

– SYSTEM NAME(*) ISOLATETIME(0) CONNFAIL(YES)

- **NAME(*)** - all systems treated equally
- **ISOLATETIME(0)** (default) – delay time prior to isolating a system in ‘status update missing’
- **CONNFAIL(YES)** – to allow SFM to handle connectivity failures vs. issuing IXC409D
- **WEIGHT(1)** – default relative importance value of a system – used in conjunction with CONNFAIL(YES) and REBUILDPERCENT

Availability
Automatic Restart Management

Can be used for automatically restarting key subsystems (IMS/DB2/CICS).

Often used for automatically restarting a customer's automation product – example for restarting CA's OPSMVS:

```
RESTART_GROUP(OPSMVS)
    TARGET_SYSTEM(*)
    FREE_CSA(600,600)
    ELEMENT(OPSMVS*)
    RESTART_ATTEMPTS(3,3600)
    RESTART_TIMEOUT(900)
    READY_TIMEOUT(900)
    TERMTYPE(ELEMTERM)
```

IBM Exploiters

- GRS lock
- JES2 checkpoint
- OPERLOG
- LOGREC
- RACF sharing
- VTAM Generic Resources
- IRD (LPAR CPU Mgmt, Dynamic Channel Path Mgmt)
- WLM Multi-system enclaves
- MQ shared queues
- RRS (Resource Recovery Services)
- Enhanced catalog Sharing (ECS)
- Batch Pipes
- VSAM RLS
- CICS TS
- DB2 Data sharing
- IMS (Shared queues, OSAM/VSAM buffer invalidates)
- IRLM Locking

RMF Reports

- Coupling Facility Activity Report - SYSRPTS(CF)
  - Structure Summary
  - Storage/Processor Summary
- XCF Activity Report – REPORTS(XCF)
  - Path Statistics
  - XCF Usage by System

RMF Reports (cont'd)

Coupling Facility Activity Report
SYSRPTS(CF)

z/OS V1R7 SYSPLEX PLEX1 DATE 04/26/2010 INTERVAL 015.00.000
RPT VERSION V1R7 RMF TIME 15.00.00 CYCLE 00.250 SECONDS

COUPLING FACILITY NAME = CF1
TOTAL SAMPLES(AVG) = 3596 (MAX) = 3596 (MIN) = 3596

COUPLING FACILITY USAGE SUMMARY

STRUCTURE SUMMARY

TYPE	STRUCTURE NAME	STATUS CHG	ALLOC SIZE	% OF CF STOR	# REQ	% OF ALL REQ	% OF CF UTIL	AVG REQ/ SEC	LST/DIR ENTRIES TOT/CUR	DATA ELEMENTS TOT/CUR	LOCK ENTRIES TOT/CUR	DIR REC/ DIR REC XI'S
LIST	IXC1K	ACTIVE	9M	1.0	14500	17.2	17.8	16.11	424	398	N/A	N/A
									1	16	N/A	N/A
	IXC16K	ACTIVE	12M	1.3	3114	3.7	11.6	3.46	1088	1072	N/A	N/A
									1	16	N/A	N/A
	OPERLOG	ACTIVE	16M	1.7	1689	2.0	2.8	1.88	13K	38K	N/A	N/A
									6066	18K	N/A	N/A
	...											
LOCK	ISGLOCK	ACTIVE	32M	3.4	42961	51.0	12.9	47.73	0	0	2097K	N/A
									0	0	9216	N/A
	...											
CACHE	SYSIGGCAS_ECS	ACTIVE	2M	0.2	0	0.0	0.0	0.00	156	152	N/A	0
									0	0	N/A	0
	...											

RMF Reports (cont'd)

Coupling Facility Activity Report
SYSRPTS(CF)

C O U P L I N G F A C I L I T Y A C T I V I T Y

PAGE 2

z/OS V1R7 SYSPLEX PLEX1 DATE 04/26/2010 INTERVAL 015.00.000
RPT VERSION V1R7 RMF TIME 15.00.00 CYCLE 00.250 SECONDS

COUPLING FACILITY NAME = CF1
TOTAL SAMPLES(AVG) = 3596 (MAX) = 3596 (MIN) = 3596

COUPLING FACILITY USAGE SUMMARY

STORAGE SUMMARY

	ALLOC SIZE	% OF CF STORAGE	------- DUMP SPACE ------- % IN USE	MAX % REQUESTED
TOTAL CF STORAGE USED BY STRUCTURES	737M	79.2		
TOTAL CF DUMP STORAGE	2M	0.2	0.0	50.0
TOTAL CF STORAGE AVAILABLE	191M	20.5		

TOTAL CF STORAGE SIZE	930M			

	ALLOC SIZE	% ALLOCATED	
TOTAL CONTROL STORAGE DEFINED			930M 79.5
TOTAL DATA STORAGE DEFINED			0K 0.0

PROCESSOR SUMMARY

COUPLING FACILITY 2098 MODEL E10 CFLEVEL 16 DYNDISP OFF

AVERAGE CF UTILIZATION (% BUSY) 0.3 LOGICAL PROCESSORS: DEFINED 1 EFFECTIVE 1.0
SHARED 0 AVG WEIGHT 0.0

RMF Reports (cont'd)

XCF Activity Report
REPORTS(XCF)

z/OS V1R7 SYSTEM ID MVS1 DATE 03/26/2010 INTERVAL 15.00.000
RPT VERSION V1R7 RMF TIME 00.00.00 CYCLE 0.250 SECONDS

TOTAL SAMPLES = 3,584 XCF PATH STATISTICS

OUTBOUND FROM MVS1

TO SYSTEM	T Y P	FROM/TO DEVICE, OR STRUCTURE	TRANSPORT CLASS	REQ OUT	AVG Q LNGTH	AVAIL	BUSY	RETRY
MVS2	S	IXC1K	TC1K	1,505	0.00	1,505	0	0
	S	IXC16K	TC16K	8,640	0.00	8,640	0	0
	S	IXC64K	TC64K	192	0.00	192	0	0
	C	043A TO 034A	TC1K	2,484	0.00	2,484	0	0
	C	0434 TO 0344	TC1K	2,737	0.00	2,737	0	0
	C	0436 TO 0346	TC1K	2,072	0.00	2,072	0	0
	C	0438 TO 0348	TC1K	1,081	0.00	1,081	0	0
MVS3	S	IXC1K	TC1K	1,221	0.00	1,221	0	0
	S	IXC16K	TC16K	157	0.00	157	0	0
	S	IXC64K	TC64K	58	0.00	58	0	0
MVS4	S	IXC1K	TC1K	658	0.00	658	0	0
	S	IXC16K	TC16K	202	0.00	202	0	0
	S	IXC64K	TC64K	0	0.00	0	0	0
TOTAL				33,961				

INBOUND TO MVS1

FROM SYSTEM	T Y P	FROM/TO DEVICE, OR STRUCTURE	REQ IN	BUFFERS UNAVAIL
MVS2	S	IXC1K	1,600	0
	S	IXC16K	1,067	0
	S	IXC64K	21	0
	C	0343 TO 0433	4,194	0
	C	0345 TO 0435	3,585	0
	C	0347 TO 0437	4,634	0
	C	0349 TO 0439	2,081	0
MVS3	S	IXC1K	1,634	0
	S	IXC16K	472	0
	S	IXC64K	20	0
MVS4	S	IXC1K	1,225	0
	S	IXC16K	871	0
	S	IXC64K	21	0
TOTAL			28,475	

RMF Reports (cont'd)

XCF Activity Report
REPORTS(XCF)

z/OS V1R7 SYSTEM ID MVS1 DATE 04/26/2010 INTERVAL 15.00.000
RPT VERSION V1R7 RMF TIME 13.30.00 CYCLE 0.250 SECONDS

XCF USAGE BY SYSTEM

REMOTE SYSTEMS — LOCAL

OUTBOUND FROM MVS1 — INBOUND TO MVS1 — MVS1

TO SYSTEM	TRANSPORT CLASS	BUFFER LENGTH	REQ OUT	BUFFER % SML	BUFFER % FIT	BUFFER % BIG	BUFFER % OVR	ALL PATHS UNAVAIL	REQ REJECT	FROM SYSTEM	REQ IN	REQ REJECT	TRANSPORT CLASS	REQ REJECT
MVS2	DEFAULT	956	0	0	0	0	0	0	0	MVS2	22,470	0	DEFAULT	0
	TC1K	956	11,458	0	100	0	0	0	0				TC1K	0
	TC16K	16,316	19,097	100	0	0	0	0	0				TC16K	0
	TC64K	62,464	62	100	0	0	0	0	0				TC64K	0
MVS3	DEFAULT	956	0	0	0	0	0	0	0	MVS3	2,857	0		
	TC1K	956	1,786	0	100	0	0	0	0					
	TC16K	16,316	772	100	0	0	0	0	0					
	TC64K	62,464	160	100	0	0	0	0	0					
MVS4	DEFAULT	956	0	0	0	0	0	0	0	MVS4	3,089	0		
	TC1K	956	1,644	0	100	0	0	0	0					
	TC16K	16,316	661	100	0	0	0	0	0					
	TC64K	62,464	0	0	0	0	0	0	0					
		TOTAL	35,640							TOTAL	28,416			

Publications

- System z10 PR/SM Planning Guide SB10-7153
- MVS Setting Up a Sysplex SA22-7625
- MVS Initialization and Tuning Reference SA22-7592
- MVS Initialization and Tuning Guide SA22-7591
- MVS System Commands SA22-7627
- z/OS Parallel Sysplex Operational Scenarios SG24-2079
- RMF Users Guide SC33-7990
- RMF Report Analysis SC33-7991
- RMF Performance Mgmt Guide
  Chapter 6 - Analyzing Sysplex Activity SC33-7992